7-7-2020

Collaborative Research: Mechanism and Target Recognition of Protein Arginine Methyltransferases (PRMTs)

Joan Hevel

Utah State University, joanie.hevel@usu.edu

Data Management Plan
Hevel (USU) & Acevedo (Miami) Groups

Products of the Research

All data generated under this NSF grant will be used towards advancing scientific understanding through publication in peer-reviewed journals and presentations at scientific meetings. The data produced in the Hevel laboratory will be utilized to characterize PRMT enzymes. The data are numerical (kinetic measurements) or are digital representations from chromatograms, gels/Westerns, protein coordinates and three-dimensional renderings. Other products include plasmids for the production of recombinant proteins. The data in the Acevedo lab include final and intermediate numerical results from theoretical calculations and software for performing simulations.

Responsibility

The principal investigator will have overall responsibility for data management over the course of the research project and will monitor compliance.

Data Format

In the Acevedo lab, data will be stored on the principal investigator's external redundant hard drive array which is hosted at the home institute and maintained by information technology personnel. Data is also backed up to cloud-based storage site Box, which has been certified and approved by the University of Miami for the storage of highly sensitive data (HIPAA/HITECH, FERPA, PCI).

A narrative of data collection in the Hevel lab is recorded in notebooks, including the salient data results. Instrument-generated data, such as gel images, spectra or chromatograms, are stored in hard copies as printouts and/or instrument outputs, in addition to electronic versions, such as ASCII codes, TIFF or JPEG, CSV files, MS Office documents, or other data-appropriate formats. All digital data is also backed up to cloud-based storage site Box drive, accessible to all lab members. .

In the Acevedo lab, data documentation will begin at the project's inception and all numeric data will be associated with the program format generated by the computational software. Data generated will come in many forms, e.g., simulation trajectories, structural data sets, analysis data. Simulation trajectories generated typically range in size from gigabytes to terabytes and are usually in .mdcrd formats. Outputs generated from the simulation programs and analysis scripts are usually stored in .out and .txt. Most numerical data are inputted in Microsoft Excel Spreadsheet format, *i.e.*, xls or xlsx. In addition to the original data format, secondary electronic data will be converted to file formats that will more likely be accessible in the long term, *i.e.*, PDF, ASCII, or TIFF.

Access to Data and Data Sharing Practices and Policies

In general, access to obtained/produced data will be made available to the public through scientific publications and poster and oral presentations. Research data that has been formally

published in a peer-reviewed journal will be made available to the public via the principal investigator's departmental website. In addition, all supplemental information for the articles, including additional graphs, tables, and information from output files, will be uploaded to the journal publisher's supplementary materials section of the website and will be made available free of charge to the public via the internet.

Data sharing will be adhered to by releasing data in response to a specific request from an interested party. To ensure the appropriate protection of privacy, confidentiality, security, and intellectual properties, provisions will be made to protect the data. These include: (1) thorough review to assess disclosure risk, (2) modifying information, if necessary, to protect confidentiality, (3) limiting access to datasets in which risk of disclosure remains high, and (4) developing an agreed upon procedure with data producers to manage disclosure risk.

Policies for Re-Use, Re-Distribution and Production of Derivatives

Links to published work will be made available on the departmental websites and through USU's institutional repository DigitalCommons@USU By depositing data on these websites, the principal investigator does not transfer copyright, but instead grants permission for the public to redisseminate the data and facilitates preservation. Disclaimers will be given on the website alerting the user to conditions regarding the use of the data in publications and products. As a general rule, users of the data or images will be required to cite the published article containing the original data in their work.

Archiving of Data

All data generated in the PI's labs are stored for at least three years beyond the end of any NSF funding period. Computational data in electronic format will be preserved for 10 years after being published in peer-reviewed journals. To protect against water and fire damage, hardcopy notebooks, and printouts of spectra or instrument outputs are stored on elevated shelving in a room with fire-stop doors. Electronic data storage devices include instruments' internal memory, external hard drives, and CD/DVD discs. As a regular practice, stored data is/will be periodically transferred to a new storage media to ensure compatibility with emerging technologies. Additionally, a periodic full back-up of all electronic data is/will be conducted and stored on an external hard-drive in a separate building, or on an external server. Samples of plasmids are archived in two different freezers in two different rooms. In the Acevedo lab, primary and secondary data in electronic format will be stored on the investigator's laboratory computers with data backup mechanisms.. Data in electronic format will be preserved for at least 10 years after being published in peer-reviewed journals. Hard copy notebooks will be retained for at least 10 years.